

Figure 13. Monthly Terminal Connect Time by Community

2.2.3 Individual Project and Community Usage

The table following shows cumulative resource usage by project during the past grant year. The entries include a summary of the operational funding sources (outside of SUMEX-supplied computing resources) for currently active projects, total CPU consumption by project (Hours), total terminal connect time by project (Hours), and average file space in use by project (Pages, 1 page = 512 computer words). These data were accumulated for each project for the months between May 1978 and April 1979. Again the well developed use of the resource by the Stanford community can be seen. It should be noted that the Stanford projects have voluntarily shifted a substantial part of their development work to non-prime time hours which is not explicitly shown in these cumulative data. It should also be noted that a significant part of the DENDRAL and MYCIN efforts, here charged to the Stanford aliquot, support development efforts dedicated to national community access to these systems. The actual demonstration and use of these programs by extramural users is charged to the national community in the "AIM USERS" category, however.

RESOURCE USE BY INDIVIDUAL PROJECT - 5/78 THROUGH 4/79

<u>NATIONAL AIM COMMUNITY</u>	CPU (Hours)	CONNECT (Hours)	FILE SPACE (Pages)
1) ACT PROJECT "Acquisition of Cognitive Procedures" John Anderson, Ph.D. Carnegie-Mellon Univ.	111.39	1497.82	2555
2) CHEM SYNTHESIS PROJECT "Simulation & Evaluation of Chemical Synthesis" W. Todd Wipke, Ph.D. U. California, Santa Cruz	370.90	5730.58	8339
3) MOD HUMAN COGN PROJECT (since 12/78) "Hierarchical Models of Human Cognition" Peter Polson, Ph.D. Walter Kintsch, Ph.D. University of Colorado	38.26	654.28	223

4)	HIGHER MENTAL FUNCTIONS "Intelligent Speech Prosthesis" Kenneth Colby, M.D. UCLA	30.80	490.29	2687
5)	INTERNIST PROJECT "DIALOG: Computer Model of Diagnostic Logic" Jack Myers, M.D. Harry Pople, Ph.D. University of Pittsburgh	196.99	2658.47	7832
6)	MISL PROJECT "Medical Information Systems Laboratory" Morton Goldberg, M.D. Bruce McCormick, Ph.D. U. Illinois, Chicago Cir.	3.50	132.47	1120
7)	PUFF/VM PROJECT "Biomedical Knowledge Engineering in Clinical Medicine" John Osborn, M.D. Inst. Medical Sciences, San Francisco Edward Feigenbaum, Ph.D. Stanford University	97.48	3351.63	2222
8)	RUTGERS PROJECT "Computers in Biomedicine" Saul Amarel, D.Sc.	30.63	868.12	10093

9)	SCP PROJECT	18.30	436.90	275
	"Simulation of Cognitive Processes"			
	James Greeno, Ph.D.			
	Alan Lesgold, Ph.D.			
	University of Pittsburgh			
10)	AIM PILOT PROJECTS			
	Psychopharm. Advisor	25.63	537.73	773
	Organ Culture	24.35	449.21	924
	Commun. Enhancement	1.83	121.71	329
	KRL Demonstrations	2.53	54.06	388
		-----	-----	----
	AIM Pilot Totals	54.34	1162.71	2414
11)	AIM Administration	14.58	461.15	5808
12)	AIM Users on Stanford Projects			
	AGE	1.17	82.22	14
	DENDRAL	44.37	860.51	1092
	MOLGEN	.20	6.99	24
	MYCIN	5.12	137.33	295
	Guest (all projects)	47.01	812.21	189
	Other	.63	27.74	144
		-----	-----	----
	AIM User Totals	98.50	1927.00	1762
		-----	-----	----
	COMMUNITY TOTALS	1065.67	19371.42	45330

<u>STANFORD COMMUNITY</u>	CPU (Hours)	CONNECT (Hours)	FILE SPACE (Pages)
1) AI HANDBOOK PROJECT Edward Feigenbaum, Ph.D.	80.69	1935.01	2021
2) DENDRAL PROJECT "Resource Related Research Computers and Chemistry" Carl Djerassi, Ph.D.	1315.63	19639.31	21517
3) AGE PROJECT "Generalization of AI Tools" Edward Feigenbaum, Ph.D.	28.76	1022.46	1344
4) HYDROID PROJECT "Distributed Processing and Problem Solving" Gio Wiederhold, Ph.D.	39.65	1725.03	789
5) MOLGEN PROJECT "Experiment Planning System for Molecular Genetics" Edward Feigenbaum, Ph.D. Laurence Kedes, M.D. Douglas Lenat, Ph.D.	384.31	6954.92	5730
Nancy Martin, Ph.D. U. New Mexico			
6) MYCIN PROJECT "Computer-based Consult. in Clin. Therapeutics" Bruce Buchanan, Ph.D. Edward Shortliffe, M.D., Ph.D.	499.07	8384.56	8687

Section 2.2.3

Individual Project and Community Usage

7) PROTEIN STRUCT MODELING "Heuristic Comp. Applied to Prot. Crystallog." Edward Feigenbaum, Ph.D.	206.48	2958.98	4392
8) RX PROJECT (since 2/79) Robert Blum, M.D. Gio Wiederhold, Ph.D.	7.57	608.94	312
9) STANFORD PILOT PROJECTS			
Genetics Applic.	104.50	1874.51	482
Quantum Chemistry	178.64	2004.44	810
Ultrasonic Imaging	5.32	130.67	85
Miscellaneous	.43	18.28	6
	-----	-----	-----
Stanford Pilot Totals	288.89	4027.90	1384
10) SU-ASSOCIATES	22.06	699.41	1557
	-----	-----	-----
COMMUNITY TOTALS	2873.11	47956.52	47733

<u>SUMEX STAFF</u>	CPU (Hours)	CONNECT (Hours)	FILE SPACE (Pages)
1) Staff	953.68	28941.65	9028
2) MAINSAIL Development	446.39	9045.69	3804
3) Staff associates, misc.	65.62	2776.72	4503
	-----	-----	-----
COMMUNITY TOTALS	1465.69	40764.06	17335

<u>SYSTEM OPERATIONS</u>	CPU (Hours)	CONNECT (Hours)	FILE SPACE (Pages)
1) Operations	1949.22	78944.64	81114
	=====	=====	=====
RESOURCE TOTALS	7353.69	187036.64	191512

2.2.4 Network Usage

The following plots show total terminal connect time per month for TYMNET and ARPANET users since initial connection. No corresponding plot is presented for the experimental TELENET connection because of frequent line configuration changes during the connection period and the short period of active use.

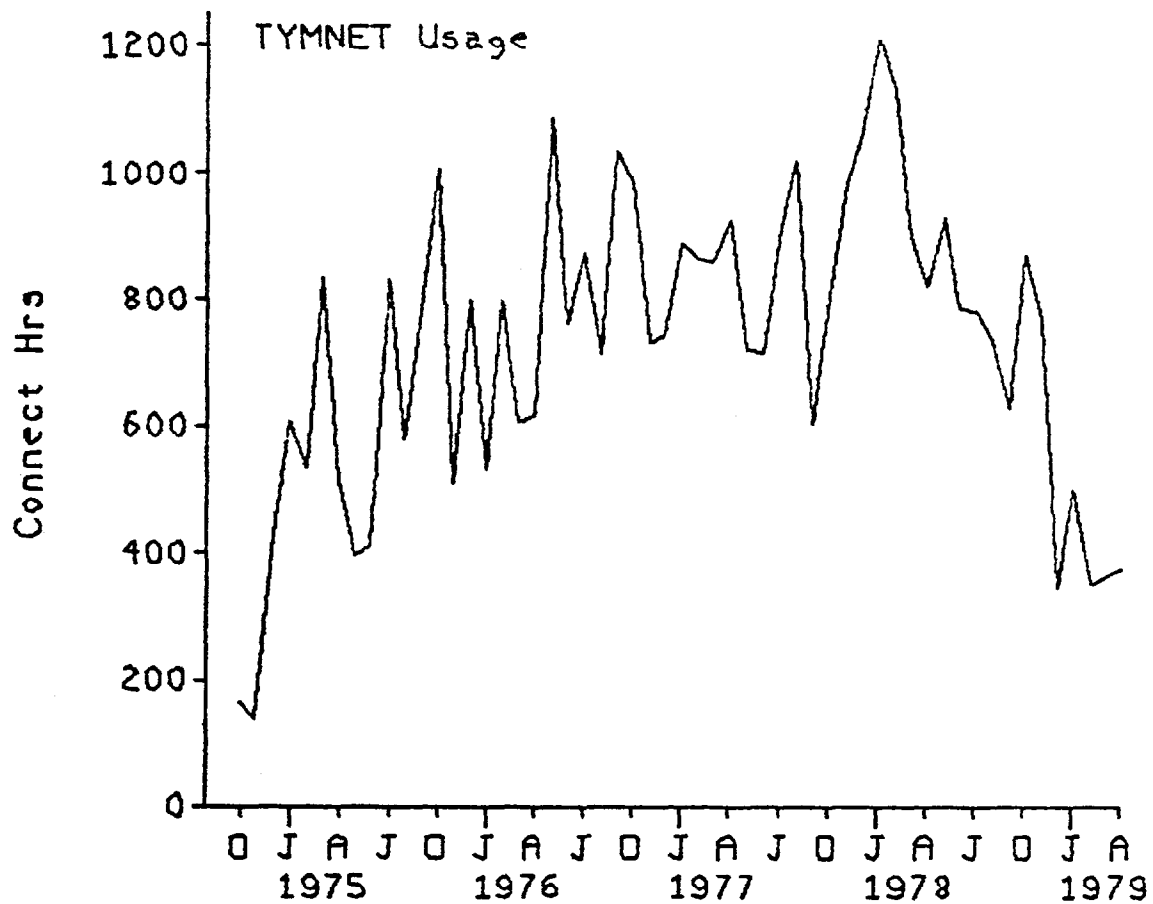


Figure 14. TYMNET Usage Data

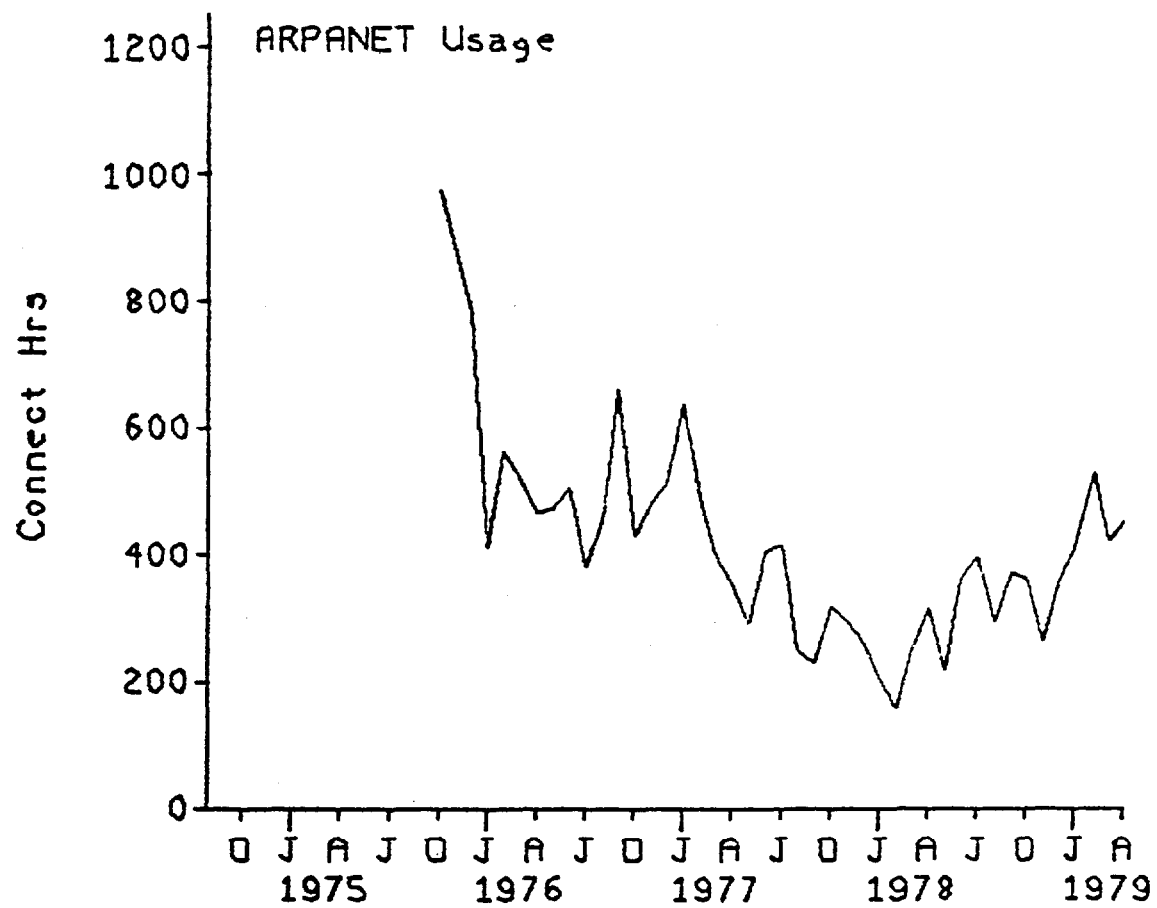


Figure 15. ARPANET Usage Data

2.3 Resource Equipment Summary

A complete inventory of resource equipment is being submitted separately along with the budget material.

2.4 Publications

The following are publications for the SUMEX staff and have included papers describing the SUMEX-AIM resource and on-going research as well as documentation of system and program developments. Publications for individual collaborating projects are detailed in their respective reports (see Section 4 on page 64).

- [1] Carhart, R.E., Johnson, S.M., Smith, D.H., Buchanan, B.G., Dromey, R.G., and Lederberg, J., Networking and a Collaborative Research Community: A Case Study Using the DENDRAL Programs, ACS Symposium Series, Number 19, Computer Networking and Chemistry, Peter Lykos (Editor), 1975.
- [2] Levinthal, E.C., Carhart, R.E., Johnson, S.M., and Lederberg, J., When Computers Talk to Computers, Industrial Research, November 1975
- [3] Wilcox, C. R., MAINSAIL - A Machine-Independent Programming System, Proceedings of the DEC Users Society, Vol. 2, No. 4, Spring 1976.
- [4] Wilcox, Clark R., The MAINSAIL Project: Developing Tools for Software Portability, Proceedings, Computer Application in Medical Care, October, 1977, pp. 76-83.
- [5] Lederberg, J. L., Digital Communications and the Conduct of Science: The New Literacy, Proc. IEEE, Vol. 66, No. 11, Nov 1978.
- [6] Wilcox, C. R., Jirak, G. A., and Dageforde, M. L., MAINSAIL - An Approach to Software Portability, in preparation.
- [7] Rindfleisch, T. C., Feigenbaum, E. A., and Lederberg, J., SUMEX-AIM - A Model for Resource Sharing and Scientific Collaboration, in preparation.

Mr. Clark Wilcox also chaired the session on "Languages for Portability" at the DECUS DECsystem10 Spring '76 Symposium.

In addition, a substantial continuing effort has gone into developing, upgrading, and extending documentation about the SUMEX-AIM resource, the SUMEX-TENEX system, the many subsystems available to users, and MAINSAIL. These efforts include a number of major documents (such as SOS, PUB, and TENEX-SAIL manuals) as well as a much larger number of document upgrades, user information and introductory notes, an ARPANET Resource Handbook entry, and policy guidelines.

3 Resource Finances

3.1 Budget Information

The budget for the SUMEX project detailing past actual costs, current year status, and estimates for the next grant year are submitted in a separate document to the NIH.

3.2 Resource Funding

The SUMEX-AIM resource is essentially wholly funded by the Biotechnology Resources Program (6). The various collaborator projects which use SUMEX are independently funded with respect to their manpower and operating expenses. They obtain from SUMEX, without charge, access to the computing and, in most cases, communications facilities in exchange for their participation in the scientific and community building goals of SUMEX.

(6) Except for participation by Stanford University in accordance with general cost-sharing and for assistance to SUMEX from other projects with overlapping aims and interests.

Collaborative Projects

4 Collaborative Project Reports

The following subsections report on the collaborative use of the SUMEX facility. Descriptions are included for the formally authorized projects within the national AIM and Stanford aliquots and the various "pilot" efforts currently under way. These project descriptions and comments are the result of a solicitation for contributions sent to each of the project Principal Investigators requesting the following information:

I. SUMMARY OF RESEARCH PROGRAM

- A. Technical goals
- B. Medical relevance and collaboration
- C. Progress summary
- D. List of relevant publications
- E. Funding support status (see below for details)

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

- A. Collaborations and medical use of programs via SUMEX
- B. Sharing and interactions with other SUMEX-AIM projects
(via workshops, resource facilities, personal contacts, etc.)
- C. Critique of resource management
(community facilitation, computer services, capacity, etc.)

III. RESEARCH PLANS (8/79 - 7/81)

- A. Long range project goals and plans
- B. Justification and requirements for continued SUMEX use
[This section will be of special importance to the Advisory Committee and is your application for continued access.]
- C. Your needs and plans for other computational resources, beyond SUMEX/AIM
- D. Recommendations for future community and resource development

We believe that the reports of the individual projects speak for themselves as rationales for participation; in any case the reports are recorded as submitted and are the responsibility of the indicated project leaders.

4.1 National AIM Projects

The following group of projects is formally approved for access to the AIM aliquot of the SUMEX-AIM resource. Their access is based on review by the AIM Advisory Group and approval by the AIM Executive Committee.

4.1.1 Acquisition of Cognitive Procedures (ACT)

Acquisition of Cognitive Procedures (ACT)

Dr. John Anderson
Carnegie-Mellon University
Pittsburgh, Pennsylvania

I. Summary of Research Program

A. Technical goals:

To develop a production system that will serve as an interpreter of the active portion of an associative network. To model a range of cognitive tasks including memory tasks, inferential reasoning, language processing, and problem solving. To develop an induction system capable of acquiring cognitive procedures with a special emphasis on language acquisition.

B. Medical relevance and collaboration:

1. The ACT model is a general model of cognition. It provides a useful model of the development of and performance of the sorts of decision making that occur in medicine.
2. The ACT model also represents basic work in AI. It is in part an attempt to develop a self-organizing intelligent system. As such it is relevant to the goal of development of intelligent artificial aids in medicine.

We have been evolving a collaborative relationship with James Greeno and Allan Lesgold at the University of Pittsburgh. They are applying ACT to modeling the acquisition of reading and problem solving skills. We have made ACT a guest system within SUMEX. ACT is currently at the state where it can be shipped to other INTERLISP facilities. We have received a number of inquiries about the ACT system. ACT is a system in a continual state of development but we periodically freeze versions of ACT which we maintain and make available to the national AI community.

C. Progress and accomplishments:

ACT provides a uniform set of theoretical mechanisms to model such aspects of human cognition as memory, inferential processes, language processing, and problem solving. ACT's knowledge base consists of two components, a propositional component and a procedural component. The propositional component is provided by an associative network encoding a set of facts known about the world. This provides the system's semantic memory. The procedural component consists of a set of productions which operate on the associative network. ACT's production system is considerably different than many of the other currently available systems (e.g., Newell's PSG). These differences have been introduced in order to create a system that will operate on an associative network and in order to accurately model certain aspects of human cognition.

A small portion of the semantic network is active at any point in time. Productions can only inspect that portion of the network which is active at that time. This restriction to the active portion of the network provides a means to focus the ACT system in a large data base of facts. Activation can spread down network paths from active nodes to activate new nodes and links. To prevent activation from growing continuously there is a dampening process which periodically deactivates all but a select few nodes. The condition of a production specifies that certain features be true of the active portion of the network. The action of a production specifies that certain changes be made to the network. Each production can be conceived of as an independent "demon." Its purpose is to see if the network configuration specified in its condition is satisfied in the active portion of memory. If it is, the production will execute and cause changes to memory. In so doing it can allow or disallow other productions which are looking for their conditions to be satisfied. Both the spread of activation and the selection of productions are parallel processes whose rates are controlled by "strengths" of network links and individual productions. An important aspect of this parallelism is that it is possible for multiple productions to be applied in a cycle. Much of the early work on the ACT system was focused on developing computational devices to reflect the operation of parallel, strength-controlled processes and working out the logic for creating functioning systems in such a computational medium.

We have successfully implemented a number of small-scale systems that model various psychological tasks in the domain of memory, language processing, and inferential reasoning. There was a larger scale project to model the language processing mechanisms of a young child. This includes implementation of a production system to analyze linguistic input, make inferences, ask and answer questions, etc.

The current research is focused on developing mechanisms for the acquisition of skills. In the framework of the ACT system this maps into acquiring new productions and modifying old productions. We have developed learning devices to enable existing productions to create new productions, to adjust the strengths of existing productions, to produce more general variants of existing productions, to produce more discriminant variants of existing productions, and to combine a number of existing productions into a single compact production. We have developed the F version of the ACT system which has these learning facilities. We have so far tested out the system in a number of small learning examples. Current goals involve applying the system to the acquisition of language skills, development of mathematical problem solving skills, and acquisition of initial programming skills.

The basic insight in this research is to model skill acquisition as an interaction between deliberate learning and automatic induction. To the extent that the teacher or the learner is able to understand the skill to be acquired, it is possible for ACT to directly create the necessary productions. However, as a fallback for less structured situations, ACT has automatic induction mechanisms that try to develop the necessary mechanisms by an intelligent trial-and-error inductive process. Much of our research has gone to identifying the heuristics used by this inductive process. Traditionally, there has been a contrast in psychology between learning with understanding and learning by trial and error. It is now clear to us that most real learning situations involve a mixture and the key to understanding skill acquisition is to understand that mixture.

One major project is the investigation of the learning of skills in Geometry. We have written several versions of a program that provides reasons, i.e. postulate names, to worked-out proofs. A number of new mechanisms were developed for this program. For instance, we developed a semantic net representation of the goal tree for problem solving. We also developed ways for the program to automatically shift from a serial search to a parallel search for relevant postulates. There were also several applications of ACT's general learning mechanisms to learn and speed up the use of postulates.

D. Current list of project publications:

- [1] Anderson, J.R. Language, Memory, and Thought. Hillsdale, N.J.: L. Erlbaum, Assoc., 1976.
- [2] Kline, P.J. & Anderson, J.R. The ACTE User's Manual, 1976.
- [3] Anderson, J.R., Kline, P. & Lewis, C. Language processing by production systems. In P. Carpenter and M. Just (Eds.). Cognitive Processes in Comprehension. L. Erlbaum Assoc., 1977.
- [4] Anderson, J.R. Induction of augmented transition networks. Cognitive Science, 1977, 125-157.
- [5] Anderson, J.R. & Kline, P. Design of a production system. Paper presented at the Workshop on Pattern-Directed Inference Systems, Hawaii, May 23-27, 1977.
- [6] Anderson, J.R. Computer simulation of a language acquisition system: A second report. In D. LaBerge and S.J. Samuels (Eds.). Perception and Comprehension. Hillsdale, N.J.: L. Erlbaum Assoc., 1978.
- [7] Anderson, J.R., Kline, P.J., & Beasley, C.M. A theory of the acquisition of cognitive skills. In G.H. Bower (Ed.). Learning and Motivation, Vol. 13. New York: Academic Press, 1979.
- [8] Anderson, J.R., Kline, P.J., & Beasley, C.M. Complex Learning. In R. Snow, P.A. Frederico, & W. Montague (Eds.). Aptitude, Learning, an Instruction: Cognitive Processes Analyses. Hillsdale, N.J.: Lawrence Erlbaum Assoc., 1979.

II. Interaction With the SUMEX-AIM Resource

A. & B. Collaborations, interactions, and sharing of programs via SUMEX.

We have received and answered many inquiries about the ACT system over the ARPANET. This involves sending documentations, papers, and copies of programs.

The most extensive collaboration has been with Greeno and Lesgold who are also on SUMEX (see the report of the Simulation of Comprehension Processes project). There is an ongoing effort to assist them in their research. Feedback from their work is helping us with system design.

We find the SUMEX-AIM workshops ideal vehicles for updating ourselves on the field and for getting to talk to colleagues about aspects of their work of importance to us.

Due to memory space problems encountered by ACT (see section III.A.2) we expect that soon we will need to make use of the smaller version of INTERLISP developed at SUMEX for use in the CONGEN program.

C. Critique of resource management.

The SUMEX-AIM resource has been well suited for the needs of our project. We have made the most extensive use of the INTERLISP facilities and the facilities for communication on the ARPANET. We have found the SUMEX personnel extremely helpful both in terms of responding to our immediate emergencies and in providing advice helpful to the long-range progress of the project. Despite the fact that we are not located at Stanford, we have not encountered any serious difficulties in using the SUMEX system; in fact, there are real advantages in being in the Eastern time zone where we can take advantage of the low load on the system during the morning hours. We have been able to get a great deal of work done during these hours and try to save our computer-intensive work for this time.

Two location changes by the ACT project (from Michigan to Yale in the summer of 1976 and from Yale to Carnegie-Mellon in the summer of 1978) have demonstrated another advantage of working on SUMEX: In both cases we were back to work on SUMEX the day after our arrival.

III. Research Plans (8/79-7/81)

A. Long-range user project goals and plans:

Our long-range goals are: (1) Continued development of the ACT system; (2) Application of the system to modeling of various cognitive processes; (3) Dissemination of the ACT system to the national AI community.

1. System Development Efficiency problems are the most serious ones currently facing the ACT system. Even, the modest-size simulations of learning we have done (about 100 productions) run out of space in INTERLISP after 200 cycles and each cycle may take almost a minute of real-time during periods of moderate system load. We are developing the capability to represent productions as compiled LISP code which should significantly improve the speed of the system and, perhaps even more important, should alleviate space problems because of INTERLISP's ability to overlay compiled code. We also hope to implement ACT in the smaller versions of INTERLISP that have been developed at SUMEX.

2. Application to Modeling Cognitive Processes. We anticipate a gradual decrease in the amount of effort that will go into system development and an increase in the amount of effort that will go into application of the system for modeling. We mentioned above the modeling efforts that we are using to assess the suitability of the ACTF system. We have long-range commitments to apply the ACT learning model to the following three topics: Acquisition of language (both first and second language acquisition); acquisition of programming skills; acquisition of problem solving skills in the domain of geometry. We find each of these topics to be considerable interest in and of themselves, but they also will serve as strong tests of the learning model. We are hopeful that the systems that are acquired by ACT will satisfy computational standards of good artificial intelligence. Therefore, in future years we would also be interested in applying the ACT model to acquisition of cognitive skills in medically related domains such as diagnosis or scientific inference. SUMEX would be an ideal location for collaboration on such a project.

We are also designing a system that will learn to give reasons to proofs. It will have the ability to use existing knowledge about such things as iteration, to accept instructions from a textbook, and to automatically become more efficient as it works on proofs. One learning mechanism we are very interested in is composition, a more general version of the transitive rule of inference used to combine productions. It promises to be interesting in its ability to change goal trees while problem solving. We will investigate it further.

3. Dissemination of the ACT project Although a guest version of ACT has been implemented, a user manual will have to be completed for this version before it is truly accessible to guests. A manual for the E version of ACT has existed for some time, but a manual for the F(learning) version of ACT is currently in preparation.

B. Justification for continued use of SUMEX:

Our goal for the ACT system is that it should serve as a ready-made "programming language" available to members of the cognitive science community for assembling psychologically-accurate simulations of a wide range of cognitive processes. Our intention and ability to provide such a resource justifies our use of the SUMEX facility. This facility is designed expressly for the purpose of developing and supporting such national AI resources and is, in this regard, clearly superior to the (otherwise outstanding) facilities we have available locally from the Carnegie-Mellon computer science department. Among the most important SUMEX advantages are the availability of INTERLISP on a machine accessible by either the ARPANET or TYMNET and the existence of a GUEST login. It appears that, at least for the time being, ACT has no hope of being a national resource unless it resides at SUMEX and, given the local unavailability of a network-accessible INTERLISP, it would even be very difficult to shift any significant portion of our development work from SUMEX to CMU.

C. Needs and plans for other computational resources

Carnegie-Mellon's plans to begin upgrading its PDP-10 hardware to emerging state-of-the-art machines (VAX, LISP machines, etc.) promises to provide a

excellent resource eventually, and we hope to have access to that resource as it develops. However, given that a considerable amount of software development will be required, a sophisticated LISP system such as INTERLISP is not likely to be available on this hardware in the near future.

D. Comments and suggestions for future resource goals:

We would, of course, be delighted if the computational capacity of the SUMEX facility could be increased. The slowness of the system at peak hours is a limiting factor although it is not grievous. This problem is perhaps less grievous for us than Stanford-based users because of our ability to use morning hours. We do not feel any urgent need for development of new software.

4.1.2 Chemical Synthesis Project (SECS)

SECS - Simulation and Evaluation of Chemical Synthesis

Principal Investigator: W. Todd Wipke
Board of Studies in Chemistry
University of California at Santa Cruz

Coworkers: (Postdoctoral Fellows) S. Krishnan, C. Buse, and M. Huber
(Graduate Students) G. Ouchi and D. Dolata
(Programmers) T. Blume, M. Toy, and M. Case

I. SUMMARY OF RESEARCH PROGRAM

A. Technical Goals.

The long range goal of this project is to develop the logical principles of molecular construction and to use these in developing practical computer programs to assist investigators in designing stereospecific syntheses of complex bio-organic molecules. Our specific goals this past year focused on basic research into representation of strategies, incorporation of automatic processing of functional group interchange, and preparing a robust version of SECS for updating the ADP network copy and prerelease to NIH and other collaborators.

B. Medical Relevance and Collaboration.

The development of new drugs and the study of how drug structure is related to biological activity depends upon the chemist's ability to synthesize new molecules as well as his ability to modify existing structures, e.g., incorporating isotopic labels or other substituents into biomolecular substrates. The Simulation and Evaluation of Chemical Synthesis (SECS) project aims at assisting the synthetic chemist in designing stereospecific syntheses of biologically important molecules. The advantages of this computer approach over normal manual approaches are many: 1) greater speed in designing a synthesis; 2) freedom from bias of past experience and past solutions; 3) thorough consideration of all possible syntheses using a more extensive library of chemical reactions than any individual person can remember; 4) greater capability of the computer to deal with the many structures which result; and 6) capability of computer to see molecules in graph theoretical sense, free from bias of 2-D projection.

The objective of using SECS in metabolism is to predict the plausible metabolites of a given xenobiotic in order that they may be analyzed for possible carcinogenicity. Metabolism research may also find this useful in the identification of metabolites in that it suggests what to look for. Finally, it seems there may even be application of this technique in problem domains where one wishes to alter molecules so certain types of metabolism will be blocked.

C. Progress and Accomplishments.

Research Environment: At the University of California, Santa Cruz, we have a GT40 and a GT46 graphics terminal connected to the SUMEX-AIM resource by 1200 baud leased lines (one leased line supported by SUMEX). We also have a TI725, TI745, CDI-1030, DIABLO 1620, and an ADM-3A terminal used over leased lines to SUMEX. UCSC has only a small IBM 370/145, a PDP-11/45 and 11/70 (the latter are limited to small student time-sharing jobs of 12 K words per user), all of which are unsuitable for this research. The SECS laboratory is located in the same building as the synthetic chemists at Santa Cruz so there is very facile interaction.

THE SECS PROGRAM is a large interactive program. On SUMEX it occupies about 150K words if not overlaid and about 68K when overlaid. SECS is generally used from a GT4X terminal, but can with less convenience be used from a teletype. In the former case, the chemist draws in the target molecule to be synthesized using the light-pen. The basic sequence then is that the program analyzes the structure for rings, functional groups, stereochemistry, etc., builds a three-dimensional model, and if appropriate also a Huckel Molecular Orbital model of the pi-systems, and finally on the basis of this knowledge, selects from a library of chemical transforms those reactions which could be used in the last step of the synthesis of this target. First the program reviews the generated precursors to see that they do not violate simple chemical rules of valence and stability, then the chemist reviews the precursors to delete those that seem uninteresting, and to select one for further processing in the same way the original target structure was processed.

Bug Fixes, Additions and Modifications: In the past year considerable effort has been devoted to the elimination of bugs and improvement of human engineering features. All bugs which had been found by us or reported by other users have been corrected. By deliberately requesting SECS to perform contradictory or ambiguous tasks, several additional bugs were uncovered and fixed. The addition of some simple routines to handle input has made it virtually impossible for the user to crash the program by giving it incorrect input. The overall result is that SECS 2.7 is by far the most robust version of the program ever produced and is the pre-release version being made available to those who request it.

SECS Users Manual: The previous SECS Users Manual (version 2.0) has been completely rewritten to include the extensive additions and modifications which have been made since the release of version 2.0. The manual provides not only operating instructions, but background information and examples to show users how best to use SECS 2.7.

Hardcopy of the Synthesis Tree: A user can now specify structures in the synthesis tree to be plotted. This can be by individual structure, the lineage of a structure, or conditions such as all structures with a priority value greater than 60 or that have been rated "GOOD". A separate program then drives a local Zeta plotter to plot the synthesis tree with structures, transform names and priorities. The user specifies the format of the tree. Trees containing thousands of structures can be plotted--the plot is simply generated in strips that are later pasted together. This facilitates sending a chemist a permanent record of the synthesis tree that can be mounted on his wall and provide guidance to his ongoing experimental project.

Alchem Library: We received a number of transforms which had originally been written by the SECS group and subsequently modified by chemists at Merck. Most of these transforms are tremendous improvements. However, some transforms, particularly those involving bond migrations, had been modified in such a way that chemically reasonable transformations could be suppressed for what are purely strategic reasons. Our philosophy has always been to keep chemistry and strategic considerations separate. The Merck-modified transforms have been included in our chemistry library. Our current focus is on strategic control, but we are correcting ALCHEM transform errors when they appear. It is hoped that as SECS is used by more sites, we will receive additional input to our current library of approximately 400 transforms.

Strategic Control: In the early days of computer synthesis, the major problems were in representing reactions so the computer could carry them out correctly. The problem has now shifted to the question of how to properly guide the program efficiently toward pathways which are not only chemically plausible, but are also synthetically significant. We refer to this guiding as strategic control. Without strategic control, SECS applies all reactions that "fit" the target, which generates one level of the synthesis tree. Although in theory the chemist could select appropriate precursors and still find many good syntheses, in practice so many precursors are generated that it is difficult to pick out the "good" precursors, it is difficult to foresee where a given precursor might ultimately lead, and it is so tiring that one doesn't explore the synthesis tree as completely as one should. Feedback from users of SECS indicates they too recognize that strategic control is a major urgent need for this research.

The problem is to control the program without introducing unnecessary bias, since freedom from bias is the computer's advantage over manual analysis. We have developed a philosophy and an implementation which we feel may solve this problem. We define strategy as a general principle which helps guide one in generating a simple synthesis. Strategies are based on symmetry, mathematical considerations of yield, economy of operations, etc. We prevent strategies from being based on any particular reaction. When a strategy is applied to a particular synthetic target molecule, it generates goals. Goals are described only in terms of molecular structural changes or features, and may not, for example, refer to reactions. Thus, strategies create goals, and both are completely independent of the reaction library.

Our list structured language continues to evolve as need for new expressions occurs. We have generalized its structure to allow for any number of machine generated goals and improved the human interface to the goals, preventing accidental recursive goals, and providing extensive help and explanation of how to create and modify goals. Much of our effort has been directed toward creating goals to save the chemist time and to assure that good goals are not accidentally overlooked.

The following paragraphs describe some of the current strategy work.

Subgoals. When a chemical transform has a high priority and seems to be able to satisfy a goal on the goal list the transform is "relevant", but still may not be "applicable" owing to some mismatch between what the transform requires and what the operand structure has. This mismatch can spawn a SUBGOAL to change the structure until this transform is applicable. The first

utilization of subgoals in SECS is for automatic functional group interchange (FGI).

The new subgoals have been expanded to encompass enough information to allow the program to continue from the point where a structural mismatch forced the initial halt. After the subgoal has been satisfied, and the FGI intermediate has been created, SECS then returns to the originating transform and proceeds with the application of that transform. After this has been done for all subgoal created intermediates, SECS then presents the chemist with the multi-step tree that is produced.

On complex molecules with large number of functional groups many subgoals are created, even when duplicates are prevented. This caused problems due to storage limitations. This problem has been partially solved by enabling SECS to estimate the likelihood of success of the subgoal originating transform before generation or application of the subgoal. This not only saves space by preventing the creation of subgoals whose creating transform will predictably fail, but also saves CPU time by eliminating the need to try to satisfy these fruitless subgoals. In test cases, from 50% to 75% of the originating transforms could be shown to predictably fail, thus saving that much space and time.

Since this process involves looking at transforms in an uncertain environment, not all failures can be predicted. Approximately 10% of the subgoals created still lead to "useless" intermediates. However, none of the eliminated subgoals would have led to "fruitful" intermediates, so the process is quite acceptable.

A Functional Group Oriented Strategy. Another machine-generated strategy based on the functional groups present in the target molecule has been implemented in the SECS program. In its present form, those transforms which utilize functional groups regarded as sensitive are favored over those which do not. The effect is to focus the attention of the program on one part of the molecule until the sensitive functional group(s) are removed or altered or until that part of the molecule is removed completely. At present, three levels of functional group sensitivity have been defined for this purpose: very sensitive, sensitive and not sensitive. The classification of a particular functional group depends on its sensitivity toward a range of reaction conditions and its "protectability".

Similarity. We have previously reported the development of an algorithm for determining the degree of similarity between two chemical structures. Although that algorithm was mathematically satisfying in that $s=1.0$ only when the two structures were identical, it was time consuming to calculate. We have now developed a second algorithm, which is more empirical, but very rapidly computed. This second algorithm has been compared with the first on many examples and is found to be quite good for finding when two structures are synthetically similar. Both algorithms take into account atom types, bond types, stereochemistry, functional groups, rings, etc. Papers describing these functions are in draft form soon to be submitted for publication.

Currently the similarity module requires a special version of SECS. We plan in the next year to incorporate this module into the standard version of SECS so that the bonds that if broken could lead to identical or similar